Research Report

Identifiability of Parameters in Item Response Models With Unconstrained Ability Distributions

Shelby J. Haberman

Research & Development

December 2005
RR-05-24

Identifiability of Parameters in Item Response Models With Unconstrained Ability Distributions

Shelby J. Haberman
ETS, Princeton, NJ

December 2005

Abstract

If a parametric model for the ability distribution is not assumed, then the customary two-parameter and three-parameter logistic models for item response analysis present identifiability problems not encountered with the Rasch model. These problems impose substantial restrictions on possible models for ability distributions.

Key words: Ability distribution, latent variable, cumulant generating function, moments

Acknowledgements

The author would like to thank Matthias von Davier and Paul Holland for helpful discussions and references.

In the two-parameter logistic (2PL) and three-parameter logistic (3PL) models commonly employed in item response analysis, parameter estimation is typically accomplished by use of marginal maximum likelihood based on an assumption of a normal ability distribution (Bock & Lieberman, 1970; Bock & Aitkin, 1981); however, attempts have been made to consider marginal estimation with less restricted ability distributions (Heinen, 1996, chap. 6). In Section 2, exploitation of results previously derived for the Rasch model (Cressie & Holland, 1983) permits a demonstration that marginal estimation with unrestricted ability distributions is problematic for the 2PL and 3PL models because the general 2PL and 3PL models place insufficient restrictions on the joint distribution of the item responses to permit identification of model parameters. In Section 4, the problems encountered with the general 2PL and 3PL models are shown not to exist in the Rasch model, and conditions are provided under which restricted versions of the 2PL and 3PL models place adequate restrictions on the joint distribution of the item responses so that estimation of model parameters can be considered.

Nonetheless, as noted in Section 4, estimation of parameters may remain impractical even in cases in which 2PL and 3PL models place adequate restrictions for parameter estimation to be possible in principle. This issue is examined in the context of latent class models.

1. General Marginal Estimation

To address the fundamental difficulty with marginal estimation with less restricted ability distributions requires some general results concerning maximum likelihood estimation for item responses. In a test with binary responses, random variables Y_{ij}, $1 \leq j \leq q$, $1 \leq i \leq n$, are observed, where $n \geq 1$ and $q \geq 2$ are integers, and Y_{ij} represents a response of examinee i on item j of a test. The possible values of Y_{ij} are 1 (correct) and 0 (incorrect). Let $\mathbf{Y}_i$, $1 \leq i \leq n$, denoted the q-dimensional vector of responses Y_{ij}, $1 \leq j \leq q$. If the examinees can be regarded as a simple random sample from an infinite population of possible examinees, then the $\mathbf{Y}_i$ are independent and identically distributed.

To characterize the distribution of $\mathbf{Y}_i$, some preliminary notation is helpful. Let J be the set of vectors of dimension q with coordinates 0 or 1, so that J has $m = 2^q$ elements and $\mathbf{Y}_i$ is in J. Let R^J be the set of arrays $\mathbf{r}$ with real coordinates $r_{\mathbf{y}}$ for $\mathbf{y}$ in J, and let S be the unit simplex in R^J, so that S consists of $\mathbf{r}$ in R^J such that $r_{\mathbf{y}} \geq 0$ for $\mathbf{y}$ in J and $\sum_{\mathbf{y}\in J} r_{\mathbf{y}} = 1$, and let S_0 be the set of $\mathbf{r}$ in S with all coordinates positive. For x real, let $\mathbf{x}$ be the member of R^J with all elements equal to x.

The distribution of $\mathbf{Y}_i$ is characterized by $\mathbf{p}$ in S, where, for $\mathbf{y}$ in J, $p_{\mathbf{y}}$ is the probability that $\mathbf{Y}_i = \mathbf{y}$.

Inferences concerning $\mathbf{p}$ may be based on the array $\mathbf{f}$ of relative frequencies, where, for $\mathbf{y}$ in J, $f_{\mathbf{y}}$ is the fraction of the examinees i with $\mathbf{Y}_i = \mathbf{y}$. The array $\mathbf{f}$ in S is then a sufficient statistic for $\mathbf{p}$. The log likelihood function ℓ then satisfies

$$\ell(\mathbf{p}) = n \sum_{\mathbf{p} \in J} f_{\mathbf{y}} \log p_{\mathbf{y}}$$

for $\mathbf{p}$ in S, where the convention $0 \log 0 = 0$ is used. For any nonempty subset T of S, $\hat{\mathbf{p}}$ is a maximum likelihood estimate of $\mathbf{p}$ for the model $M(T)$ that $\mathbf{p}$ is in T if $\hat{\mathbf{p}}$ is in T and $\ell(\hat{\mathbf{p}})$ is the supremum $\ell(T)$ of $\ell(\mathbf{p})$ for $\mathbf{p}$ in T.

It is well-known that, for the unrestricted model $M(S)$, the unique maximum likelihood estimate of $\mathbf{p}$ is $\mathbf{f}$. If $\mathbf{f}$ is in T for a subset T of S, then $\mathbf{f}$ is also the unique maximum likelihood estimate of $\mathbf{p}$ for model $M(T)$, for

$$\ell(\mathbf{f}) = \ell(S) \geq \ell(T)$$

and

$$\ell(\mathbf{f}) \leq \ell(T).$$

Thus the maximum $\ell(T)$ of the log likelihood for model $M(T)$ is the same as the maximum $\ell(S)$ of the log likelihood for model $M(S)$. The log likelihood ratio test statistic

$$L^2 = 2n[\ell(S) - \ell(T)] = 0.$$

In this fashion, no evidence exists to discriminate between models $M(T)$ and $M(S)$ even if T is a proper subset of S.

In large samples, it is a simple matter to find a condition under which, as the sample size n approaches ∞, the probability approaches 1 that $\mathbf{f}$ is the unique maximum likelihood estimate. Let the interior $\mathrm{ir}(T)$ of T relative to S be the union of all sets $O \cap S \subset T$ such that O is an open subset of R^J. Let model $M(T)$ be said to be locally unrestricted if $\mathrm{ir}(T)$ is nonempty, and let the model be locally unrestricted at $\mathbf{p}$ in T if $\mathbf{p}$ is in $\mathrm{ir}(T)$. The model is said to be locally restricted if it is not locally unrestricted. If model $M(T)$ is locally restricted at $\mathbf{p}$ in $\mathrm{ir}(T)$, then the weak law of large numbers implies that, as the sample size n approaches ∞, the probability approaches

1 that $\mathbf{f}$ is in $\mathrm{ir}(T)$ and $\hat{\mathbf{p}} = \mathbf{f}$ (Cramér, 1946, p. 254). Thus the probability approaches 1 that the test statistic $L^2 = 0$. In this fashion, it is clearly undesirable for a model to be locally unrestricted.

General Models for Item Responses With One-Dimensional Ability

Many apparently reasonable models for item responses are locally unrestricted. This problem is examined in this section in terms of one-parameter logistic (1PL), two-parameter logistic (2PL), and three-parameter logistic (3PL) models (Hambleton, Swaminathan, & Rogers, 1991, ch. 2). To develop a general framework for the discussion, consider the general one-dimensional model for item responses in which an ability parameter has one dimension. To each examinee i, associate an unobserved random variable θ_i that represents the ability of that examinee. The local independence assumption is made that, for each examinee i, the responses Y_{ij}, $1 \le j \le q$, are conditionally independent given θ_i. It is also assumed that the pairs $(\mathbf{Y}_i, \theta_i)$, $1 \le i \le n$, are independent and identically distributed. This latter assumption is consistent with the previous assumption that the $\mathbf{Y}_i$ are independent and identically distributed.

The common distribution function of the ability parameter θ_i is denoted by F. Associated with each item j is a nondecreasing item characteristic curve (ICC) P_j, $0 < P_j < 1$, such that, for each real θ and each item j, $1 \le j \le q$, and each examinee i, $1 \le i \le n$, $P_j(\theta)$ is the probability that $Y_{ij} = 1$ given that $\theta_i = \theta$. Let

$$Q_j = 1 - P_j \tag{1}$$

be the ICC of $1 - Y_{ij}$, and let the item logit function λ_j be

$$\lambda_j = \log(P_j/Q_j) \tag{2}$$

(Holland, 1990), so that

$$P_j = [1 + \exp(-\lambda_j)]^{-1} \tag{3}$$

and

$$Q_j = [1 + \exp(\lambda_j)]^{-1}. \tag{4}$$

Let $\boldsymbol{\lambda}$ be the q-dimensional function with coordinates λ_j for $1 \le j \le q$. For q-dimensional vectors $\mathbf{a}$ and $\mathbf{b}$ with respective coordinates a_j and b_j for $1 \le j \le q$, let

$$\mathbf{a}'\mathbf{b} = \sum_{j=1}^{q} a_j b_j.$$

Under the item response model, $\mathbf{p}$ is in S_0, and

$$p_{\mathbf{y}} = \int \left[\prod_{j=1}^{q} P_j^{y_j} Q_j^{1-y_j}\right] dF, \ \mathbf{y} \in J, \tag{5}$$

(Cressie & Holland, 1983), so that the item characteristic curves P_j, $1 \leq j \leq q$, and the distribution F determine the common joint distribution of the $\mathbf{Y}_i$. Equivalently, if

$$V = \prod_{j=1}^{q} [1 + \exp(\lambda_j)]^{-1}, \tag{6}$$

then the following variant on the Dutch identity holds:

$$p_{\mathbf{y}} = \int V \exp(\mathbf{y}'\boldsymbol{\lambda}) dF. \tag{7}$$

(Holland, 1990).

Let $\mathcal{P}$ be the set of q-dimensional functions $\mathbf{P}$ on the real line with coordinates P_j, $1 \leq j \leq q$, that are strictly increasing real functions on the real line with values in $(0, 1)$. Let Λ be the set of q-dimensional functions $\boldsymbol{\lambda}$ with coordinates λ_j, $1 \leq j \leq q$, that are strictly increasing real functions on the real line. Let $\mathcal{F}$ be the set of real functions that are distribution functions of real random variables. The assumption that the general one-dimensional model holds states in effect that $\mathbf{p}$ is in the subset S_m of all $\mathbf{p}$ in S_0 such that (5) holds for some $\mathbf{P}$ in $\mathcal{P}$ and some F in $\mathcal{F}$ (Cramér, 1946, p. 57). The set S_m is a proper subset of S_0 (Holland & Rosenbaum, 1986). The set S_m can also be defined to be the set of $\mathbf{p}$ in S_0 such that (6) and (7) hold for some $\boldsymbol{\lambda}$ in Λ and some F in $\mathcal{F}$.

In an item response model with a one-dimensional ability parameter, $\mathbf{p}$ is assumed to belong to a nonempty subset T of S_m. In this section, some common item response models are examined to determine whether they are locally unrestricted. The results summarized are obtained from Theorem 5 and from dimensional analysis.

Rasch Models

In a general Rasch (1PL) model, $\boldsymbol{\lambda}$ is in the set Λ_1 of $\boldsymbol{\lambda}$ in Λ such that

$$\boldsymbol{\lambda}(\theta) = (a\theta)\mathbf{1} - \boldsymbol{\gamma} \tag{8}$$

for some real $a > 0$ and some $\boldsymbol{\gamma}$ in R^q with coordinates γ_j, $1 \leq j \leq q$. The common item discrimination is a and the item difficulty of item j is $\beta_j = \gamma_j/a$. No restriction is made on the ability distribution. Thus $\mathbf{p}$ is in the set S_1 of $\mathbf{p}$ that satisfy (6) and (7) for F in $\mathcal{F}$ and for $\boldsymbol{\lambda}$ in Λ_1. Provided that $q > 2$, the model is locally restricted. Otherwise, the model is locally unrestricted.

A somewhat less general version of the Rasch model is the normal Rasch model that requires that F is the cumulative normal distribution function Φ (Bock & Lieberman, 1970; Bock & Aitkin, 1981). In this model, $\mathbf{p}$ is in the set S_{1n} of $\mathbf{p}$ that satisfy (6) and (7) for $F = \Phi$ and $\boldsymbol{\lambda}$ in Λ_1. As in the case of the general Rasch model, the normal Rasch model is locally restricted for $q > 2$ and locally unrestricted for $q = 2$.

In a latent-class Rasch model for a given vector $\boldsymbol{\tau}$ of distinct ability levels τ_k, $1 \le k \le K$, $K \ge 2$ (Heinen, 1996), it is assumed that the Rasch model holds for F in the set $\mathcal{F}_{\boldsymbol{\tau}}$ of distribution functions of random variables that only have values τ_k, $1 \le k \le K$. The corresponding set $S_{1\boldsymbol{\tau}}$ consists of $\mathbf{p}$ in S_m that satisfy (6) and (7) for some F in $\mathcal{F}_{\boldsymbol{\tau}}$ and some $\boldsymbol{\lambda}$ in Λ_1. For $q > 2$, the model is locally restricted. For $q = 2$, the model is locally unrestricted.

2PL Models

In a general 2PL model, $\boldsymbol{\lambda}$ is in the set Λ_2 of $\boldsymbol{\lambda}$ in Λ such that

$$\boldsymbol{\lambda}(\theta) = \theta\mathbf{a} - \boldsymbol{\gamma} \tag{9}$$

for some $\mathbf{a}$ in R^q with positive coordinates a_j and some $\boldsymbol{\gamma}$ in R^q. Thus a_j is the item discrimination and $\beta_j = \gamma_j/a_j$ is the item difficulty for item j. In the case of no restriction on the ability distribution, $\mathbf{p}$ is in the set S_2 of $\mathbf{p}$ that satisfy (6) and (7) for F in $\mathcal{F}$ and for some $\boldsymbol{\lambda}$ in Λ_2. Rather remarkably, the simple change from the constant item discrimination in the Rasch model to variable item discrimination in the 2PL model results in a model that is always locally unrestricted.

The normal 2PL model requires that F is the cumulative normal distribution function Φ. In this model, $\mathbf{p}$ is in the set S_{2n} of $\mathbf{p}$ that satisfy (6) and (7) for $F = \Phi$ and for $\boldsymbol{\lambda}$ in Λ_2. The normal 2PL model is locally restricted for $q > 2$ and locally unrestricted for $q = 2$.

In a latent-class 2PL model for a given vector $\boldsymbol{\tau}$ of distinct ability levels τ_k, $1 \le k \le K$, $K \ge 2$, it is assumed that the 2PL model holds for F in the set $\mathcal{F}_{\boldsymbol{\tau}}$. The corresponding set $S_{2\boldsymbol{\tau}}$ consists of $\mathbf{p}$ in S_m that satisfy (6) and (7) for some F in $\mathcal{F}_{\boldsymbol{\tau}}$ and for some $\boldsymbol{\lambda}$ in Λ_2. In this case, local restriction occurs if $2q + K < 2^q$, and the model is locally unrestricted if $q \le 3$. As discussed in Section 4, the theoretical existence of local restriction does not necessarily ensure that parameter estimation is really practical.

3PL Models

In a general 3PL model, it is assumed that $\mathbf{p}$ is in the set S_3 of $\mathbf{p}$ in S_m such that (1) and (5) hold for some F in $\mathcal{F}$ and some $\mathbf{P}$ in the set $\mathcal{P}_3$ of $\mathbf{P}$ in $\mathcal{P}$ such that, for real θ,

$$P_j(\theta) = c_j + (1 - c_j)[1 + \exp(-a_j\theta + \gamma_j)]^{-1},\ 1 \le j \le q, \tag{10}$$

for some real $a_j > 0$, c_j in $[0, 1)$, and γ_j, $1 \le j \le q$. Here c_j is the item guessing parameter, a_j is the item discrimination, and $\beta_j = \gamma_j/a_j$ is the item difficulty for item j. Clearly $S_1 \subset S_2 \subset S_3 \subset S_m$, so that the general 3PL model is locally unrestricted.

The normal 3PL model requires that F is the cumulative normal distribution function Φ. In this model, $\mathbf{p}$ is in the set S_{3n} of $\mathbf{p}$ that satisfy (1) and (5) for $F = \Phi$ and for $\mathbf{P}$ in $\mathcal{P}_3$. The normal 3PL model is locally restricted for $q > 3$ and locally unrestricted for $q \le 3$.

In a latent-class 3PL model for a given vector $\boldsymbol{\tau}$ of distinct ability levels τ_k, $1 \le k \le K$, $K \ge 2$, it is assumed that the 3PL model holds for F in the set $\mathcal{F}_{\boldsymbol{\tau}}$. The corresponding set $S_{3\boldsymbol{\tau}}$ consists of $\mathbf{p}$ in S_m that satisfy (1) and (5) for some F in $\mathcal{F}_{\boldsymbol{\tau}}$ and for some $\mathbf{P}$ in $\mathcal{P}_3$. In this case, local restriction occurs if $3q + K < 2^q$, and the model is locally unrestricted if $q \le 3$. Once again, the theoretical existence of local restriction does not necessarily ensure that parameter estimation is really practical.

2. The General 2PL and 3PL Cases

As already noted, the general 2PL and 3PL models are locally unrestricted. Proof relies on the following theorem.

Theorem 1 *Let $\boldsymbol{\lambda}$ be in Λ, and let (1) hold. Let T be a subset of S_m such that $\mathbf{p}$ is in T if (7) holds for some F in $\mathcal{F}$. Let $\mathbf{c}$ be the function from the real line R to R^J with coordinates*

$$c_{\mathbf{y}} = \exp(\mathbf{y}'\boldsymbol{\lambda}),\ \mathbf{y} \in J.$$

Let the $\mathbf{c}(\theta)$, θ real, span R^J. Then model $M(T)$ is locally unrestricted.

Proof. Consider distinct θ_k, $1 \le k \le m$, such that $\mathbf{c}(\theta_k)$, $1 \le k \le m$, spans R^J. Let A_m be the unit simplex in R^m, so that an m-dimensional vector $\mathbf{a}$ is in A_m if the coordinates a_k are nonnegative for $1 \le k \le m$ and if $\sum_{k=1}^{m} a_k = 1$. Let $\mathbf{a}$ be in A_m, and let F be the distribution function of a random variable that assigns probability a_k to θ_k for $1 \le k \le m$. If

$$\mathbf{p} = \sum_{k=1}^{m} a_k V(\theta_k)\mathbf{c}(\tau_k),$$

then (7) holds. Thus T includes $\mathbf{p}$. Because $\mathbf{a}$ is arbitrary, T includes the simplex D with vertices $V(\tau_k)\mathbf{c}(\tau_k)$, $1 \le k \le m$. To demonstrate that $\mathrm{ir}(T)$ is nonempty, it suffices to show that $\mathrm{ir}(D)$ is nonempty. To do so, it suffices in turn to demonstrate that

$$\sum_{k=1}^{m} a_k V(\tau_k)\mathbf{c}(\tau_k) = 0 \tag{11}$$

and

$$\sum_{k=1}^{m} a_k = 0 \tag{12}$$

only if $a_k = 0$ for $1 \le k \le m$ (Rockafellar, 1970, pp. 6, 13). By assumption, (11) implies that $a_k V(\tau_k)$ is 0 for each k, so that $a_k = 0$ for $1 \le k \le m$.

The simplest application of the theorem is to the 2PL model. Consider the following result.

Theorem 2 *The general 2PL model is locally unrestricted.*

Proof. Consider a_j, $1 \le j \le q$, such that

$$s_{\mathbf{y}} = \sum_{j=1}^{q} a_j y_j$$

has m distinct values for $\mathbf{y}$ in J. Let τ_k, $1 \le k \le m$, be distinct real numbers. Consider arbitrary real γ_j, $\le j \le q$, and let (9) and (3) hold. Let

$$t_{\mathbf{y}} = \sum_{j=1}^{q} \gamma_j y_j.$$

In Theorem 1,

$$c_{\mathbf{y}}(\tau_k) = \exp(\tau_k s_{\mathbf{y}}) \exp(-t_{\mathbf{y}}).$$

The determinant of an m by m matrix with coordinates $\exp(b_i d_k)$, $1 \le i \le m$, $1 \le k \le m$, is positive definite if b_i is strictly increasing in i and d_k is strictly increasing in k (Karlin & Studden, 1966, pp. 9–10). By elementary linear algebra, it follows that the $\mathbf{c}(\tau_k)$, $1 \le k \le m$, must be linearly independent. Because R^J has dimension $m = 2^q$, it follows that the $\mathbf{c}(\tau_k)$, $1 \le k \le m$, span R^J, so that the conditions of Theorem 1 are satisfied.

Because $S_2 \subset S_3 \subset S_m$, it follows that $\mathrm{ir}(S_2) \subset \mathrm{ir}(S_3) \subset \mathrm{ir}(S_m)$. Thus Theorem 2 implies that the general 3PL model $M(S_3)$ is locally unrestricted, as is the general one-dimensional model $M(S_m)$.

Examination of the proof of Theorem 1 reveals a further problem with the 2PL and 3PL models. In the 2PL case, if $\mathbf{a}$ is chosen as in the proof of Theorem 2, $\boldsymbol{\gamma}$ is in R^q, and $\mathbf{p}$ satisfies (6), (7), and (9), then open neighborhoods O of $\mathbf{a}$ and O' of $\boldsymbol{\gamma}$ exist such that, if $\mathbf{b}$ is in O and $\boldsymbol{\delta}$ is in O', then a distribution function G exists such that

$$\boldsymbol{\mu} = \theta\mathbf{b} - \boldsymbol{\delta},$$

$$W = \prod_{j=1}^{q}[1 + \exp(-\mu_j)]^{-1},$$

and

$$p_{\mathbf{y}} = \int W \exp(\mathbf{y}'\boldsymbol{\mu})dG.$$

Thus the model parameters a_j and γ_j are not estimable in the general 2PL model. Similar comments apply to estimation of a_j, c_j, and γ_j in the general 3PL model.

3. Dimension Theory and Local Restriction

In many cases, for a nonempty subset T of the simplex S, a determination that $M(T)$ is locally restricted is based on a determination of the topological dimension of T. The topological dimension $\dim(T)$ of T may be defined in terms of finite open coverings. Here a finite class $\mathcal{C}$ of nonempty open subsets of R^J is a finite open covering of T if each $\mathbf{x}$ in T is included in some C in $\mathcal{C}$. A finite open cover $\mathcal{D}$ of T is a refinement of $\mathcal{C}$ if to each C in $\mathcal{C}$ corresponds a D in $\mathcal{D}$ such that $D \subset C$. The finite open cover $\mathcal{D}$ has integer order $k \ge 0$ if distinct sets D_i, $1 \le i \le k+1$, in $\mathcal{D}$ exist such that $\cap_{i=1}^{k+1} D_i$ is nonempty and if no distinct sets D_i, $1 \le i \le k+2$, in $\mathcal{D}$ exist such that $\cap_{i=1}^{k+2} D_i$ is nonempty. The topological dimension $\dim(T)$ of T is the smallest integer $k \ge 0$ such that every finite open cover $\mathcal{C}$ of T has a refinement $\mathcal{D}$ of order k (Hurewicz & Wallman, 1941, pp. 5, 54, 56).

As evident from the following known results, the topological dimension as defined here does have properties intuitively expected of a definition of dimension.

Theorem 3 *If U is a nonempty subset of T and T is a subset of S, then* $\dim(U) \leq \dim(T)$ *(Hurewicz & Wallman, 1941, p. 26).*

Theorem 4 *The simplex S has topological dimension* $2^q - 1$ *(Hurewicz & Wallman, 1941, p. 46).*

Theorem 5 *If T is a nonempty subset of S, then* $\dim(T) = 2^q - 1$ *if, and only if,* $\mathrm{ir}(T)$ *is nonempty (Hurewicz & Wallman, 1941, p. 46).*

An alternative version of Theorem 5 is that $M(T)$ is locally unrestricted if, and only if, $\dim(T) = 2^q - 1$, and $M(T)$ is locally restricted if, and only if, $\dim(T) < 2^q - 1$.

Theorem 6 *If T is a nonempty subset of S, $r \geq 1$ is an integer, O is a subset of R^r with nonempty interior, $\mathbf{g}$ is a continuous one-to-one function from O onto T, and $\mathbf{g}^{-1}$ is also continuous, then T has dimension r (Hurewicz & Wallman, 1941, p. 46)*

Under the conditions of Theorem 6, an r-dimensional parameter $\boldsymbol{\eta}$ in O is uniquely defined by the equation $\mathbf{p} = \mathbf{g}(\boldsymbol{\eta})$.

Theorem 7 *If T is a nonempty subset of S, if $r \geq 1$ is an integer, if O is the union of a countable number of closed and bounded nonempty convex subsets of R^r, and $\mathbf{g}$ is a continuously differentiable function from O onto T, then* $\dim(T) \leq r$

Proof. The conclusion follows from standard results concerning Hausdorff dimension (Falconer, 1990, pp. 29–30) given the relationship of Hausdorff dimension to topological dimension (Hurewicz & Wallman, 1941, p. 104).

In Theorem 7, the condition on O holds if $O = A \cap B$ for nonempty convex subsets A and B of R^r such that A is closed and B is open (Haberman, 1996, p. 180). In all examples considered in this report, the conditions on O are satisfied.

Theorem 8 *Under the conditions of Theorem 7, let $\mathbf{g}$ have coordinates $g_{\mathbf{y}}$ for $\mathbf{y}$ in J, and let the gradient of $g_{\mathbf{y}}$ be $\nabla g_{\mathbf{y}}$. Let U be the subset of $\boldsymbol{\eta}$ in the interior of O such that the vectors $\nabla g_{\mathbf{y}}(\boldsymbol{\eta})$, $\mathbf{y}$ in J, span R^r. If U is nonempty, then* $\dim(T) = r$.

Proof. For $\boldsymbol{\eta}$ in U, there exists a subset K of J with r elements such that the $\nabla g_{\mathbf{y}}(\boldsymbol{\eta})$ are linearly independent for $\mathbf{y}$ in K. The inverse-mapping theorem then implies that a nonempty open subset N of O and a nonempty subset Z of T exist such that the restriction of $\mathbf{g}$ to N is a one-to-one function onto B with a continuous inverse (Loomis & Sternberg, 1968, p. 167). Given Theorems 3, 6, and 7, both B and T have dimension r.

Theorem 8 has significant impact on use of maximum likelihood. Let the conditions of Theorem 8 hold, and let $T \subset S_0$. Define $\boldsymbol{\eta}$, N, and Z as in the proof of Theorem 8, and let $\mathbf{p} = \mathbf{g}(\boldsymbol{\eta})$. Consider the nonsingular information matrix

$$\mathbf{I} = \sum_{\mathbf{y} \in J} [g_{\mathbf{y}}(\boldsymbol{\eta})]^{-1} \nabla g_{\mathbf{y}}(\boldsymbol{\eta}) [\nabla g_{\mathbf{y}}(\boldsymbol{\eta})]'.$$

Let $\hat{\boldsymbol{\eta}}$ in N be a function of $\mathbf{f}$ such that $\mathbf{g}(\hat{\boldsymbol{\eta}}) = \hat{\mathbf{p}}$ whenever $\hat{\mathbf{p}}$ is a maximum-likelihood estimate of $\mathbf{p}$ under the model $M(Z)$, so that $\hat{\boldsymbol{\eta}}$ is the maximum-likelihood estimate of $\boldsymbol{\eta}$. Then $n^{1/2}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})$ converges in distribution to a multivariate normal random vector with mean $\mathbf{0}$ and covariance matrix $\mathbf{I}^{-1}$ (Birch, 1964). If $\mathbf{g}$ is a one-to-one function with a continuous inverse, then N can be defined to be the interior of O and $\hat{\boldsymbol{\eta}}$ can be defined so that $\mathbf{g}(\hat{\boldsymbol{\eta}}) = \hat{\mathbf{p}}$ whenever $\hat{\mathbf{p}}$ is a maximum-likelihood estimate of $\mathbf{p}$ under model $M(T)$.

In Theorem 8, the condition that $\nabla g_{\mathbf{y}}(\boldsymbol{\eta})$, $\mathbf{y}$ in J, spans R^r is equivalent to the condition that no $\mathbf{b}$ in R^r exist such that $\mathbf{b} \neq \mathbf{0}$ and $\mathbf{b}'\nabla g_{\mathbf{y}}(\boldsymbol{\eta}) = 0$ for all $\mathbf{y}$ in J.

The Rasch Model

Given these preliminaries, it is readily shown that the general Rasch model is locally restricted if, and only if, $q \leq 2$. Consider the following theorem.

Theorem 9 *The topological dimension* $\dim(S_1) = 2q - 1$*, so that the general Rasch model is locally restricted if, and only if,* $q > 2$.

Proof. Let $u_{\mathbf{y}} = \sum_{j=1}^{q} y_j$ for $\mathbf{y}$ in J, so that $u_{\mathbf{y}}$ is a nonnegative integer not greater than q. Let $\mathbf{g}$ be the real function on R^{2q-1} such that

$$g_{\mathbf{y}}(\mathbf{x}) = c(\mathbf{x}) \exp\left(\sum_{j=1}^{2q-1} x_j t_{j\mathbf{y}}\right), \tag{13}$$

$$c(\mathbf{x}) = \left[\sum_{\mathbf{y} \in J} \exp\left(\sum_{j=1}^{2q-1} x_j t_{j\mathbf{y}}\right)\right]^{-1}, \tag{14}$$

and

$$t_{j\mathbf{y}} = \begin{cases} 1, & 1 \le u_{\mathbf{y}} = j \le q, \\ 0, & u_{\mathbf{y}} \ne j, 1 \le j \le q, \\ 1, & y_{j+1-q} = 1, q+1 \le j \le 2q-1, \\ 0, & y_{j+1-q} \ne 1, q+1 \le j \le 2q-1. \end{cases} \tag{15}$$

Let S_{1e} be the image of $\mathbf{g}$. Then $S_1 \subset S_{1e}$ (Tjur, 1982; Cressie & Holland, 1983; Haberman, 2004). The function $\mathbf{g}$ from R^{2q-1} onto S_{1e} is continuously differentiable and has a continuous inverse $\mathbf{g}^{-1}$ (Haberman, 1973), so that Theorem 6 implies that $\dim(S_{1e}) = 2q-1$. Because $S_1 \subset S_{1e}$, Theorem 3 implies that $\dim(S_1) \le 2q-1$.

To discuss the relationship of S_1 and S_{1e} requires consideration of two determinants. Let $\mathbf{A}_k(\mathbf{x})$ be defined for $\mathbf{x}$ in R^{2q-1} and for $0 \le k \le q$ in the following fashion. If k is even, then let $d = k/2$ and let $\mathbf{A}_k(\mathbf{x})$ be the $d+1$ by $d+1$ matrix with row i and column j equal to $\exp(x_{i+j-2})$ for integers i and j from 1 to $d+1$. If k is odd, then let $d = (q+1)/2$, and let $\mathbf{A}_k(\mathbf{x})$ be the d by d matrix with row i and column j equal to $\exp(x_{i+j-1})$ for integers i and j from 1 to d. Let N be the set of $\mathbf{x}$ in R^{2q-1} such that the determinants of $\mathbf{A}_q(\mathbf{x})$ and $\mathbf{A}_{q-1}(\mathbf{x})$ are positive. The set N is nonempty (Karlin & Studden, 1966, pp. 38, 171). Because a determinant of a matrix is a continuous function of the elements of the matrix, N is a nonempty open subset of R^{2q-1}. If $\mathbf{x}$ is in N, then $\mathbf{g}(\mathbf{x})$ is in S_1 (Cressie & Holland, 1983; Lindsay, Clogg, & Grego, 1991; Haberman, 2004). By Theorems 3 and 6, $\dim(S_1) = 2q-1$. Because $2q-1 < 2^q - 1$ if, and only if, $q > 2$, it follows that from Theorem 5 that the general Rasch model is locally restricted if, and only if, $q > 2$.

In Theorem 9, local restriction does not eliminate all problems of parameter estimation. One has $\mathbf{g}(\boldsymbol{\eta}) = \mathbf{p}$ for a unique $\boldsymbol{\eta}$ in S_1 if $\mathbf{p}$ is in S_1, and the inverse of $\mathbf{g}$ is continuous; however, the parameter $\boldsymbol{\eta}$ is not a simple quantity. Given (6), (7), and (8), $\eta_j = \gamma_{j+1-q} - \gamma_1$ for $q+1 \le j \le 2q-1$, so that contrasts between the γ_j are readily estimated, but

$$\exp(\eta_j) = \frac{\int V(\theta)\exp(aj\theta)dF(\theta)}{\int V(\theta)dF(\theta)}$$

for $1 \le j \le q$, so that the item discrimination a and ability distribution F are not determined by the general Rasch model (Lindsay et al., 1991).

Given Theorem 9, it follows that the normal Rasch model and the latent-class Rasch model are both locally restricted if $q > 2$. A bit more can be said in terms of parameter estimation given dimensional analysis. Consider the following results.

Theorem 10 *In the normal Rasch model,* $\dim(S_{1n}) = q+1$, *and* $M(S_{1n})$ *is locally unrestricted only if* $q = 2$.

Proof. Let O be the subset of R^{q+1} of vectors $\mathbf{x}$ with $x_{q+1} > 0$. For $\mathbf{x}$ in O, let

$$h(\mathbf{x}, \theta) = \prod_{j=1}^{q} [1 + \exp(x_{q+1}\theta x_{q+1} - x_j)]^{-1},$$

and let

$$v_{\mathbf{y}}(\mathbf{x}) = \sum_{j=1}^{q} y_j x_j.$$

Define $u_{\mathbf{y}}$ as in the proof of Theorem 9. Let $\mathbf{g}$ be the function from O onto S_{1n} with coordinates $g_{\mathbf{y}}$, $\mathbf{y}$ in J, such that

$$g_{\mathbf{y}}(\mathbf{x}) = \exp[-v_{\mathbf{y}}(\mathbf{x})] \int h(\mathbf{x}, \theta) \exp(x_{q+1} u_{\mathbf{y}} \theta) \phi(\theta) d\theta.$$

Under the normal Rasch model, (6), (7), (8), and $F = \Phi$ imply that $\mathbf{p} = \mathbf{g}(\boldsymbol{\eta})$ if $\eta_j = \gamma_j$ for $1 \le j \le q$ and $\eta_{q+1} = a$.

Given general results for exponential families with incomplete data, it is readily verified that $\mathbf{g}$ is continuously differentiable (Sundberg, 1974). Let

$$d_j(\mathbf{x}, \theta) = [1 + \exp(x_{q+1}\theta - x_j)]^{-1}$$

for $1 \le j \le q$, and let

$$d_+(\mathbf{x}, \theta) = \sum_{j=1}^{q} d_j(\mathbf{x}, \theta).$$

The gradient $\nabla g_{\mathbf{y}}(\mathbf{x})$ of $g_{\mathbf{y}}$ at $\mathbf{x}$ has elements $g_{j\mathbf{y}}(\mathbf{x})$, $1 \le j \le q+1$, such that $g_{j\mathbf{y}}(\mathbf{x})$ is

$$-\exp[-v_{\mathbf{y}}(\mathbf{x})] \int [y_j - d_j(\mathbf{x}, \theta)] h(\mathbf{x}, \theta) \exp(x_{q+1} u_{\mathbf{y}} \theta) \phi(\theta) d\theta$$

for $1 \le j \le q$ and

$$\exp[-v_{\mathbf{y}}(\mathbf{x})] \int [u_{\mathbf{y}} - d_+(\mathbf{x}, \theta)] h(\mathbf{x}, \theta) \exp(x_{q+1} u_{\mathbf{y}} \theta) \phi(\theta) d\theta$$

for $j = q+1$. By Theorem 7, $\dim(S_{1n}) \le q+1$.

Let $\mathbf{b}$ in R^{q+1} satisfy

$$\mathbf{b}' \nabla g_{\mathbf{y}}(\mathbf{x}) = 0 \tag{16}$$

for all $\mathbf{y}$ in J. Let

$$c_{\mathbf{y}}(\mathbf{x},\theta)=\sum_{j=1}^{q}(b_j+b_{q+1}\theta)[y_j-d_j(\mathbf{x},\theta)].$$

Then (16) can only hold if

$$\int c(\mathbf{x},\theta)h(\mathbf{x},\theta)\exp(x_{q+1}u_{\mathbf{y}}\theta)\phi(\theta)d\theta=0. \tag{17}$$

Consider $\mathbf{b}$ in R^{q+1} such that (16) holds. Let $\mathbf{y}$ in J have $y_1=0$ and all other coordinates 0. For $2\le j\le q$, let $\mathbf{z}$ in J have $z_j=1$ and all other coordinates 0. Then $c_{\mathbf{y}}(\mathbf{x})-c_{\mathbf{z}}(\mathbf{x})$ reduces to the constant b_1-b_j, so that (17) implies that $b_j=b_1$. Thus

$$\int[b_1+\theta b_{q+1}][k-d_+(\mathbf{x},\theta)]h(\mathbf{x},\theta)\exp(kx_{q+1}\theta)\phi(\theta)d\theta=0$$

for each integer k from 0 to q. The case of $k=0$ yields

$$b_1=-b_{q+1}\frac{\int\theta d_+(\mathbf{x},\theta)h(\mathbf{x},\theta)\phi(\theta)d\theta}{\int d_+(\mathbf{x},\theta)h(\mathbf{x},\theta)\phi(\theta)d\theta}.$$

Given simple use of a change of variable formula, the case of $k=q$ yields

$$b_1=b_{q+1}\frac{\int\theta d_+(\mathbf{r},\theta)h(\mathbf{r},\theta)\phi(\theta)d\theta}{\int d_+(\mathbf{r},\theta)h(\mathbf{r},\theta)\phi(\theta)d\theta}$$

if $r_j=-x_j$ for $1\le j\le q$ and $r_{q+1}=x_{q+1}$.

An analysis of signs provides a simple means to verify if $b_1=b_{q+1}=0$. For any $\mathbf{x}$ in O,

$$\int d_+(\mathbf{x},\theta)h(\mathbf{x},\theta)\phi(\theta)d\theta>0$$

and

$$\int d_+(\mathbf{r},\theta)h(\mathbf{r},\theta)\phi(\theta)d\theta>0.$$

It follows that $\mathbf{b}=\mathbf{0}$ if

$$\int\theta d_+(\mathbf{x},\theta)h(\mathbf{x},\theta)\phi(\theta)d\theta<0 \tag{18}$$

and

$$\int\theta d_+(\mathbf{r},\theta)h(\mathbf{r},\theta)\phi(\theta)d\theta<0. \tag{19}$$

The inequalities (18) and (19) hold if $\mathbf{x}$ has coordinates $x_j=0$ for $1\le j\le q$, for $\mathbf{x}$ and $\mathbf{r}$ are the same, and

$$h(\mathbf{x},\theta)d_+(\mathbf{x},\theta)-h(\mathbf{x},-\theta)d_+(\mathbf{x},-\theta)$$

is equal to

$$h(\mathbf{x},\theta)d_+(\mathbf{x},\theta)[1-\exp((q-1)x_{q+1}\theta)]<0$$

for $\theta>0$. Thus

$$\begin{aligned}
&\int \theta d_+(\mathbf{x},\theta)h(\mathbf{x},\theta)\phi(\theta)d\theta \\
=\ & \int_0^\infty \theta[\exp(x_{q+1}\theta)-\exp(qx_{q+1}\theta)]d_+(\mathbf{x},\theta)h(\mathbf{x},\theta)\phi(\theta)d\theta \\
<\ & 0.
\end{aligned}$$

It follows that $\mathbf{x}$ in O can be selected so that $\mathbf{b}$ must be $\mathbf{0}$ if (16) holds for all $\mathbf{y}$ in J. By Theorem 8, $\dim(S_{1n})=q+1$, so that Theorem 5 implies that $M(S_{1n})$ is then locally unrestricted if, and only if, $q+1=2^q-1$. Because $q+1=2^q-1$ only holds if $q=2$, $M(S_{1n})$ is only locally unrestricted if $q=2$.

Given Theorem 9, the latent-class Rasch model is obviously locally restricted if $q>2$, for $S_{1\boldsymbol{\tau}}\subset S_1$. The following result shows that the latent-class Rasch model is locally unrestricted if $q\le 3$.

Theorem 11 *If $q\le 3$, then $S_{1\boldsymbol{\tau}}=S_1$, so that the latent-class Rasch model $M(S_{1\boldsymbol{\tau}})$ is locally unrestricted for $q=2$.*

Proof. If $\mathbf{p}$ is in S_1, then there exist $a>0$, $\boldsymbol{\gamma}$ in R^q, $\rho_1<\rho_2$, and nonnegative π_1 and π_2 such that $\pi_1+\pi_2=1$, (6), (7), and (8) hold for F the distribution function of a random variable that assigns probability π_k to ρ_k for k equal 1 or 2 (Karlin & Studden, 1966; Lindsay et al., 1991). Let G be the distribution function of a random variable that assigns probability π_k to τ_k for k equal 1 or 2. Let

$$b=a(\rho_2-\rho_1)/(\tau_2-\tau_2),$$

and let $\boldsymbol{\delta}$ in R^q satisfy

$$\boldsymbol{\delta}=\boldsymbol{\gamma}+a\frac{\tau_2\rho_1-\tau_1\rho_2}{\tau_1-\tau_2}\mathbf{1}.$$

Let

$$\boldsymbol{\mu}(\theta)=\theta\mathbf{b}-\boldsymbol{\delta}$$

for θ real, and let

$$W=\prod_{j=1}^{q}[1+\exp(\mu_j)]^{-1}.$$

Then

$$p_{\mathbf{y}} = \int W \exp(\mathbf{y}'\boldsymbol{\mu}) dG,$$

so that $\mathbf{p}$ is in $S_{1\boldsymbol{\tau}}$. It follows that $S_{1\boldsymbol{\tau}}$ and S_1 are the same, so that the conclusions of the theorem follows from Theorem 9.

In general, an argument similar to that in Theorem 10 may be used to demonstrate that $\dim(S_{1\boldsymbol{\tau}}) \leq q + K$. Consider the following theorem.

Theorem 12 *In the latent-class Rasch model,* $\dim(S_{1\boldsymbol{\tau}}) \leq q + K$.

Proof. Let O be the set of $\mathbf{x}$ in R^{q+K} such that $x_{q+1} > 0$, $x_j \geq 0$ for $j > q + 1$ and $\sum_{j=q+2}^{q+K} x_j \leq 1$. Let h be the real function such that

$$h(\mathbf{x}, \theta) = \prod_{j=1}^{q} [1 + \exp(x_{q+1}\theta - x_j)]^{-1}$$

for $\mathbf{x}$ in O and real θ. Define $u_{\mathbf{y}}$ as in the proof of Theorem 9. Let $\mathbf{g}$ be the function from O onto $S_{1\boldsymbol{\tau}}$ such that the coordinate $g_{\mathbf{y}}$, $\mathbf{y}$ in J, satisfies

$$\begin{aligned} g_{\mathbf{y}}(\mathbf{x}) &= \sum_{k=1}^{K-1} x_{q+1+k} h(\mathbf{x}, \tau_k) \exp(x_{q+1} u_{\mathbf{y}} \tau_k) \\ &\quad + \left(1 - \sum_{k=1}^{K-1} x_{q+1+k}\right) h(\mathbf{x}, \tau_K) \exp(x_{q+1} u_{\mathbf{y}} \tau_K). \end{aligned}$$

Note that $\mathbf{p} = \mathbf{g}(\boldsymbol{\eta})$ if (6), (7), and (8) hold, $\sum_{k=1}^{K} \pi_k = 1$, F is the distribution function for a random variable that equals τ_k with probability π_k for $1 \leq k \leq K$, $\eta_j = \gamma_j$ for $1 \leq j \leq q$, $\eta_{q+1} = a$, and $\eta_{q+k+1} = \pi_k$ for $1 \leq k \leq K - 1$. It is easily verified that $\mathbf{g}$ is continuously differentiable. By Theorem 7, $\dim(S_{1\boldsymbol{\tau}})$ does not exceed $q + K$.

The upper bound in Theorem 12 is not necessarily achieved. For example, Theorem 11 implies that $\dim(S_{1\boldsymbol{\tau}}) = 3 < q + K$ if $q = 2$. On the other hand, Theorem 11 also implies that the upper bound of $q + K = 5$ is achieved if $q = 3$ and $K = 2$. In general, because $\dim(S_1) = 2q - 1$, $\dim(S_{1\boldsymbol{\tau}})$ cannot exceed $2q - 1$. Thus standard conditions for estimation of the parameters γ_j, $1 \leq j \leq q$, a, and τ_k, $1 \leq k \leq K - 1$ cannot hold if $K > q - 1$.

Restricted 2PL Models

Restricted 2PL models may be locally restricted. The required arguments are quite similar to those used for restricted Rasch models. In the case of the normal 2PL model, the relationship $S_{1n} \subset S_{2n}$ and Theorem 10 clearly implies that $\dim(S_{2n}) = 3$ and $M(S_{2n})$ is locally unrestricted if $q = 2$. For $q > 2$, $\dim(S_{2n}) = 2q + 1$ and $M(S_{2n})$ is locally restricted.

Given Theorem 11 and the relationship $S_{1\boldsymbol{\tau}} \subset S_{2\boldsymbol{\tau}}$, it is clearly true that the latent-class 2PL model is locally unrestricted for $q = 2$. For $q \leq 3$, $2q + 1 \leq \dim(S_{2\boldsymbol{\tau}}) \leq 2q + K - 1$, so that Theorem 5 implies that the model is locally restricted if $2q + K < 2^q$ and locally unrestricted if $q = 3$. The case of $q = 3$ can also be considered by use of equations used in latent-class analysis that involve determinants and eigenvalues (Madansky, 1960). This argument relies on continuity properties of eigenvalues and eigenvectors (Wilkinson, 1965, chaps. 1–2).

In principle, the latent-class 2PL model is locally restricted as long as $K < 2^q - 2q$; however, as evident in Section 4, it should not be concluded that use of a large value of K is wise.

Restricted 3PL Models

Restricted 3PL models also may be locally restricted. Arguments are again similar to those used in Theorems 10 and 12. In the normal 3PL model, the relationship $S_{2n} \subset S_{3n} \subset S$ and Theorem 3 imply that $\dim(S_{3n}) = 2^q - 1$ and $M(S_{3n})$ is locally unrestricted for $q \leq 3$. For $q > 3$, $M(S_{3n})$ is locally restricted and $\dim(S_{3n}) = 3q$.

In the case of latent-class 3PL models, the relationship $S_{2\boldsymbol{\tau}} \subset S_{3\boldsymbol{\tau}}$ implies that $M(S_{3\boldsymbol{\tau}})$ is locally unrestricted if $q \leq 3$. For $q > 3$, $\dim(S_{3\boldsymbol{\tau}}) \leq 3q + K - 1$. If $K = 2$, then $S_{3\boldsymbol{\tau}} = S_{2\boldsymbol{\tau}}$, so that $\dim(S_{3\boldsymbol{\tau}}) = \dim(S_{2\boldsymbol{\tau}}) = 2q + 1$ for $q \geq 3$, and $M(S_{3\boldsymbol{\tau}})$ is locally restricted for $q > 3$. To verify the identity of $S_{2\boldsymbol{\tau}}$ and $S_{3\boldsymbol{\tau}}$, consider the following theorem.

Theorem 13 *If $K = 2$, then $S_{2\boldsymbol{\tau}} = S_{3\boldsymbol{\tau}}$.*

Proof. Because $S_{2\boldsymbol{\tau}} \subset S_{3\boldsymbol{\tau}}$, it suffices to show that $S_{3\boldsymbol{\tau}}$ is in $S_{2\boldsymbol{\tau}}$. If $\mathbf{p}$ is in $S_{3\boldsymbol{\tau}}$, then (1), (5), and (10) hold for some $a_j > 0$, real γ_j, and c_j in $[0, 1)$ and some distribution function F such that, for some nonnegative π_1 and π_2 with sum $\pi_1 + \pi_2 = 1$, F is the distribution function of a random variable that assigns probability π_k to τ_k for $1 \leq k \leq K = 2$. Define $\mathbf{b}$ and $\boldsymbol{\gamma}$ in R^q by solution of the simultaneous equations

$$b_j\tau_k - \delta_j = \log[P_j(\tau_k)/Q_j(\tau_k)],\ 1 \leq k \leq 2,$$

for $1 \leq j \leq q$. Let $\boldsymbol{\mu}(\theta) = \theta\mathbf{b} - \boldsymbol{\delta}$. Then

$$P_j(\tau_k) = \{1 + \exp[-\mu_j(\tau_k)]\}^{-1}$$

and

$$Q_j(\tau_k) = \{1 + \exp[\mu_j(\tau_k)]\}^{-1}$$

for $1 \leq j \leq q$, so that $\mathbf{p}$ is in $S_{2\boldsymbol{\tau}}$.

4. Parameter Estimation for Latent-Class Models for Item Responses

Even though simple conditions are available to ensure that latent-class 2PL and 3PL models are locally restricted, these conditions do not imply that parameter estimation is readily accomplished. Some problems reflect fundamental problems of parameter identification that occur when the upper bound for the topological dimension is not achieved. For example, in a latent-class 3PL model with $K = 2$, it is not possible to identify a_j, c_j, and γ_j, as is evident from Theorem 13.

More complex problems involve cases in which the conditions of Theorem 7 are satisfied for $S_{2\boldsymbol{\tau}}$ or $S_{3\boldsymbol{\tau}}$ for some function $\mathbf{g}$ but the information matrix $\mathbf{I}$ is nearly singular. This situation has adverse effects on numerical algorithms for computation of maximum-likelihood estimates and adverse effects on the accuracy of parameter estimates. Such difficulties have been noted previously (Heinen, 1996), and they are relevant both with the EM algorithm (Dempster, Laird, & Rubin, 1977) and with the stabilized Newton-Raphson algorithm (Haberman, 1988). In particular, the problems of parameter estimation for latent-class 2PL and 3PL models has practical consequences for item calibration in the National Assessment of Educational Progress (NAEP) based on the Parscale/NAEP program option to employ a latent-class 3PL model with 41 latent classes.

The importance of the problem is examined in this section by use of the latent-class 2PL model. It should be emphasized that problems will be substantially more severe in the 3PL case. For analysis in this section, the parametrization uses the subset O of R^r, $r = 2q + K - 1$, such that $\mathbf{x}$ is in O if $x_j > 0$ for $q + 1 \leq j \leq 2q$, $x_j \geq 0$ for $j > 2q$, and $\sum_{j=2q+1}^{r} x_j \leq 1$. For $\mathbf{x}$ in O, θ real, and $\mathbf{y}$ in J, let

$$h(\mathbf{x}, \theta) = \prod_{j=1}^{q} [1 + \exp(x_{q+j}\theta - x_j)]^{-1},$$

$$v_{\mathbf{y}}(\mathbf{x}) = \sum_{j=1}^{q} x_j y_j,$$

$$t_{\mathbf{y}}(\mathbf{x}) = \sum_{j=1}^{q} x_{q+j} y_j,$$

and

$$c_{\mathbf{y}}(\mathbf{x}, \theta) = \exp[-v_{\mathbf{y}}(\mathbf{x})] h(\mathbf{x}, \theta) \exp[\theta t_{\mathbf{y}}(\mathbf{x})].$$

Let the function $\mathbf{g}$ from O onto $S_{2\boldsymbol{\tau}}$ be defined so that, for $\mathbf{y}$ in J, $\mathbf{g}(\mathbf{x})$ has coordinate

$$g_{\mathbf{y}}(\mathbf{x}) = \sum_{k=1}^{K} x_{2q+k} c_{\mathbf{y}}(\mathbf{x}, \tau_k)$$

for $\mathbf{y}$ in J, where x_{2q+K} denotes $1 - \sum_{j=2q+1}^{r} x_j$. If, for some real $a_j > 0$ and real γ_j, $1 \le j \le q$, (6), (7), and (9) hold for the distribution function F of a random variable that is equal to τ_k with probability $\pi_k \ge 0$, $1 \le k \le K$, where $\sum_{k=1}^{K} \pi_k = 1$, then $\mathbf{p} = \mathbf{g}(\boldsymbol{\eta})$ for $\boldsymbol{\eta}$ in O such that $x_j = \gamma_j$ for $1 \le j \le q$, $x_{q+j} = a_j$ for $1 \le j \le q$, and $x_{2q+j} = \pi_j$ for $1 \le j \le K-1$.

Let

$$d_j(\mathbf{x}, \theta) = [1 + \exp(-x_{q+j}\theta + x_j)]^{-1}$$

for $1 \le j \le q$. Then the partial derivatives of $g_{\mathbf{y}}$ are the continuous functions $g_{j\mathbf{y}}$ with values $g_{j\mathbf{y}}(\mathbf{x})$ of

$$-\sum_{k=1}^{K} x_{2q+k} [y_j - d_j(\mathbf{x}, \tau_k)] c_{\mathbf{y}}(\mathbf{x}, \tau_k)$$

for $1 \le j \le q$,

$$\sum_{k=1}^{K} x_{2q+k} \tau_k [y_j - d_j(\mathbf{x}, \tau_k)] c_{\mathbf{y}}(\mathbf{x}, \tau_k)$$

for $q + 1 \le j \le 2q$, and

$$c_{\mathbf{y}}(\mathbf{x}, \tau_k) - c_{\mathbf{y}}(\mathbf{x}, \tau_K)$$

for $2q + 1 \le j \le r$. Thus all conditions in Theorem 7 are satisfied. In Theorem 8, it is often the case that U is indeed nonempty and $\boldsymbol{\eta}$ is in U, but the matrix $\mathbf{I}$ is so close to singular that $\boldsymbol{\eta}$ cannot be accurately estimated by use of a reasonable sample size.

The problem of near singularity of $\mathbf{I}$ arises primarily from the behavior of the partial derivatives $g_{j\mathbf{y}}$ for $j > 2q$ that are associated with the probabilities π_k, $1 \le k \le K-1$. Recall from elementary linear algebra that the smallest eigenvalue of $\mathbf{I}$ is the minimum of $\mathbf{b}'\mathbf{I}\mathbf{b}$ for $\mathbf{b}$ in R^r such that $\mathbf{b}'\mathbf{b} = 1$. Let $\mathbf{z}_{\mathbf{y}}$ denote the K-dimensional vector with coordinates $c_{\mathbf{y}}(\boldsymbol{\eta}, \tau_k)$ for $1 \le k \le K$. Then this minimum is certainly no greater than the minimum of

$$\Psi(\mathbf{b}) = \sum_{\mathbf{y} \in J} p_{\mathbf{y}}^{-1} (\mathbf{b}'\mathbf{z}_{\mathbf{y}})^2$$

for $\mathbf{b}$ in R^K such that $\sum_{k=1}^{K-1} b_k^2 = 1$ and $\sum_{k=1}^{K} b_k = 0$.

For fixed $\boldsymbol{\eta}$, $c_{\mathbf{y}}(\boldsymbol{\eta}, \tau)$ is infinitely differentiable in τ. Let $\bar{v}$ be the average of the $v_{\mathbf{y}}$ for $\mathbf{y}$ in J, and let $\bar{\tau}$ be the average of the τ_k for $1 \leq k \leq K$. For each integer $r \geq 1$, Taylor's theorem can be used to show that $\mathbf{z}_{\mathbf{y}}$ has a polynomial approximation $\boldsymbol{\nu}_{\mathbf{y}}$ with coordinate k equal to

$$\exp(t_{\mathbf{y}}) \sum_{u=0}^{r} \sum_{s=0}^{u} w_{us} (\tau_k - \bar{\tau})^u (v_{\mathbf{y}} - \bar{v})^s.$$

If $(r+1)/2 \leq K$, then the b_k can be selected so that $\mathbf{b}'\boldsymbol{\nu}_{\mathbf{y}}$ is 0 for each $\mathbf{y}$ in J. Thus

$$\Psi(\mathbf{b}) = \sum_{\mathbf{y} \in J} p_{\mathbf{y}}^{-1} [\mathbf{b}'(\mathbf{z}_{\mathbf{y}} - \boldsymbol{\nu}_{\mathbf{y}})]^2.$$

As K increases and the τ_k are all in a fixed finite interval, the argument implies that the minimum of $\Psi(\mathbf{b})$ should approach 0 rather rapidly. Thus it is reasonable to expect that $\mathbf{I}$ will be nearly singular even for K somewhat less than $2^q - 2q$.

This argument suggests that the problem of near singularity is less severe for a given number of latent categories K if q is larger, for the variability of $v_{\mathbf{Y}}$ is then increased. On the other hand, the argument is also relevant in the Rasch model case of a_i constant, so that the use of a model with a finite number of values of θ_i does not necessarily lead to well-identified parameters even in the Rasch model unless K is relatively small.

For some understanding of the issue, consider a case in which $q = 11$; the item difficulties are $\beta_j = -3 + 3(j-1)/5$ and the item discriminations are $a_j = 0.5 + (j-1)/10$ for $1 \leq j \leq q$ and the examinee ability values and respective probabilities are $\tau_k = k - 3$ and $\pi_k = 0.2$ for $1 \leq k \leq 5$. Recall that $\gamma_j = a_j \beta_j$ for $1 \leq j \leq q$. In this case, the largest element of $\mathbf{I}^{-1}$ is 341.5, so that each parameter has an asymptotic standard deviation less than 0.1 if the sample size is about 34,000. If K is increased to 10, the π_k are each 0.1, and $\tau_k = -1.8 + 0.4(k-1)$ for $1 \leq k \leq 10$, then the largest element of the inverse of the information matrix of $\mathbf{Y}_i$ is 8,302,862.9, so that comparable accuracy of parameter estimates requires about 830,000,000 observations. Thus in the second case, it is unreasonable to expect satisfactory performance from maximum likelihood. In the former case, some hope exists. The most extreme problems involve diagonal elements of the information matrix that correspond to the latent probabilities π_k; however, it should be emphasized that other parameter estimates are also affected. The smallest value of a diagonal element of $\mathbf{I}^{-1}$ is only 0.09 for the case of five latent classes, while the corresponding figure for 10 latent classes is 72.3.

The problems considered here are not solved if the a_j are all constant, as in the Rasch model. For the case just examined with $K = 10$, change all a_j to 1. Then the maximum value of an element of $\mathbf{I}^{-1}$ is 1,734,948.5.

Even if the Rasch model is assumed to hold with each $a_j = 1$, problems persist. Define $\mathbf{I}$ based on the function $\mathbf{g}$ in the proof of Theorem 12. The inverse information matrix for this case has a largest diagonal element of size 1,734,210.6. The difficulties encountered do disappear if the latent probabilities π_k are known. For the original 2PL case of $K = 10$, the maximum element of the inverse information matrix for the remaining parameters γ_j and a_j is only 77.9. Comparable results are achieved for the normal 2PL model. In this case, the maximum element of the inverse information matrix is 22.1.

It is somewhat difficult to characterize precisely when near singularity occurs given that q, β_j, a_j, K, π_k, and τ_k all have impact; however, it certainly should not be assumed that use of the 2PL model with latent classes will be satisfactory without regard to the choice of K and τ_k. In typical situations, it should also be understood that results for the 3PL model are likely to be even worse.

5. Conclusions

Unlike in the case of the Rasch model, the 2PL and 3PL models do not provide a simple approach for parameter estimation in which a parametric model for the ability distribution is not assumed. In addition, simple approaches based on latent classes can be very unsatisfactory unless the number of latent classes is rather small. Thus attempts to use 2PL and 3PL models with more general ability distributions than the standard normal distribution require quite careful work, and steps must be taken to verify that parameters are determined with reasonable accuracy.

References

Birch, M. W. (1964). A new proof of the Pearson-Fisher theorem. *The Annals of Mathematical Statistics*, *35*, 817–824.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, *46*, 443–459.

Bock, R. D., & Lieberman, M. (1970). Fitting a response model for n dichotomously scored items. *Psychometrika*, *35*, 179–197.

Cramér, H. (1946). *Mathematical methods in statistics.* Princeton, NJ:: Princeton University Press.

Cressie, N., & Holland, P. W. (1983). Characterizing the manifest probabilities of latent trait models. *Psychometrika*, *48*, 129–141.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B*, *39*, 1–38.

Falconer, K. (1990). *Fractal geometry: Mathematical foundations and applications.* New York: John Wiley.

Haberman, S. J. (1973). Log-linear models for frequency data: Sufficient statistics and likelihood equations. *The Annals of Statistics*, *1*, 617–632.

Haberman, S. J. (1988). A stabilized Newton-Raphson algorithm for log-linear models for frequency tables derived by indirect observation. *Sociological Methodology*, *18*, 193–211.

Haberman, S. J. (1996). *Advanced statistics: Volume I. Description of populations.* New York: Springer-Verlag.

Haberman, S. J. (2004). *Maximum likelihood for the Rasch model for binary responses* (ETS RR-04-20). Princeton, NJ.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory.* Newbury Park, CA.

Heinen, T. (1996). *Latent class and discrete latent trait models.* Thousand Oaks, CA: Sage.

Holland, P. W. (1990). The Dutch identity: A new tool for the study of item response models. *Psychometrika*, *55*, 5–18.

Holland, P. W., & Rosenbaum, P. R. (1986). Conditional association and unidimensionality in monotone latent variable models. *The Annals of Statistics*, *14*, 1523–1543.

Hurewicz, W., & Wallman, H. (1941). *Dimension theory.* Princeton, NJ: Princeton University Press.

Karlin, S., & Studden, W. J. (1966). *Tchebycheff systems: With applications in analysis and statistics.* New York: Interscience.

Lindsay, B., Clogg, C. C., & Grego, J. (1991). Semiparametric estimation in the Rasch model and related exponential response models, including a simple latent class model for item analysis. *Journal of the American Statistical Association*, *86*, 96–107.

Loomis, L. H., & Sternberg, S. (1968). *Advanced calculus.* Reading, MA: Addison-Wesley.

Madansky, A. (1960). Determinantal methods in latent class analysis. *Psychometrika*, *25*, 183–198.

Rockafellar, R. T. (1970). *Convex analysis.* Princeton, NJ: Princeton University Press.

Sundberg, R. (1974). Maximum likelihood theory for incomplete data from an exponential family. *Scandinavian Journal of Statistics*, *1*, 49-58.

Tjur, T. (1982). A connection between Rasch's item analysis model and a multiplicative poisson model. *Scandinavian Journal of Statistics*, *9*, 23–30.

Wilkinson, J. J. (1965). *The algebraic eigenvalue problem.* Oxford, England: Clarendon Press.